Scientific Research

Optimizing Forest Sampling by Using Lagrange Multipliers

Kyriaki Kitikidou

Department of Forestry and Management of the Environment and Natural Resources,
Democritus University of Thrace, Orestiada, Greece
Email: kkitikid@fmenr.duth.gr

ABSTRACT

In two-phase sampling, or double sampling, from a population with size N we take one, relatively large, sample size n. From this relatively large sample we take a small sub-sample size m, which usually costs more per sample unit than the first one. In double sampling with regression estimators, the sample of the first phase n is used for the estimation of the average of an auxiliary variable X, which should be strongly related to the main variable Y (which is estimated from the sub-sample m). Sampling optimization can be achieved by minimizing cost C with fixed $\mathrm{var}\,\overline{Y}$, or by finding a minimum $\mathrm{var}\,\overline{Y}$ for fixed C. In this paper we optimize sampling with use of Lagrange multipliers, either by minimizing variance of Y and having predetermined cost, or by minimizing cost and having predetermined variance of Y.

Keywords: Forest Inventories; Lagrange Multipliers; Optimization, Sampling

1. Introduction

All decision-making requires information. In forestry, this information is acquired by means of forest inventories, systems for measuring the extent, quantity and condition of forests [1]. More specifically, the purpose of forest inventories is to estimate means and totals for measures of forest characteristics over a defined area. Such characteristics include the volume of the growing stock, the area of a certain type of forest and nowadays also measures concerned with forest biodiversity, e.g. the volume of dead wood or vegetation.

The main method used in forest inventories in the 19th century was complete enumeration, but it was soon noted that there was a possibility to reduce costs by using representative samples [2]. Sampling-based methods were used in forestry a century before the mathematical foundations of sampling techniques were described [3-9]. In this paper we attempt to optimize sampling with use of Lagrange multipliers, either by minimizing variance of the forest variable we are interested in and having fixed cost, or by minimizing cost and having fixed variance of the variable in question.

2. The Method of Lagrange Multipliers

Lagrange multipliers is a method of evaluating maxima or minima of a function of possibly several variables, subject to one or more constraints [10]. This method, which is due to Joseph Louis de Lagrange (1736-1813),

is used to optimize a real-valued function $f(x_1, x_2, \cdots, x_n)$, where x_1, x_2, $\cdots$, x_n are subject to m ($<n$) equality constraints of the form

$$
\begin{aligned}
g_1(x_1, x_2, \cdots, x_n) &= 0 \\
g_2(x_1, x_2, \cdots, x_n) &= 0 \\
&\vdots \\
g_m(x_1, x_2, \cdots, x_n) &= 0,
\end{aligned}
\tag{1}
$$

where g_1, g_2, $\cdots$, g_n are differentiable functions.

This determination of the stationary points in this constrained optimization problem is done by first considering the function

$$
F(x) = f(x) + \sum_{j=1}^{m} \lambda_j g_j(x)
\tag{2}
$$

where $x = (x_1, x_2, \cdots, x_n)'$ and λ_1, λ_2, $\cdots$, λ_m are scalars called Lagrange multipliers. By differentiating (2) with respect to x_1, x_2, $\cdots$, x_n and equating the partial derivatives to zero we obtain

$$
\frac{\partial F}{\partial x_i} = \frac{\partial f}{\partial x_i} + \sum_{j=1}^{m} \frac{\partial g_i}{\partial x_i} = 0, \, i = 1, 2, \cdots, n.
\tag{3}
$$

Equations (1) and (3) consist of $m + n$ unknowns, namely, x_1, x_2, $\cdots$, x_n; λ_1, λ_2, $\cdots$, λ_m. The solutions for x_1, x_2, $\cdots$, x_n determine the locations of the stationary points. The following argument explains why this is the case. Suppose that in Equation (1) we can solve for m x_i's,

for example, x_1, x_2, $\cdots$, x_n, in terms of the remaining $n - m$ variables. By Implicit Function Theorem (see Appendix 1), this is possible whenever

$$\frac{\partial(g_1, g_2, \cdots, g_m)}{\partial(x_1, x_2, \cdots, x_m)} \neq 0 . \qquad (4)$$

In this case, we can write

$$
\begin{aligned}
x_1 &= h_1(x_{m+1}, x_{m+2}, \cdots, x_n) \\
x_2 &= h_2(x_{m+1}, x_{m+2}, \cdots, x_n) \\
&\vdots \\
x_m &= h_m(x_{m+1}, x_{m+2}, \cdots, x_n).
\end{aligned}
\qquad (5)
$$

Thus $f(x)$ is a function of only $n - m$ variables, namely, x_{m+1}, x_{m+2}, $\cdots$, x_n. If the partial derivatives of f with respect to these variables exist and if f has a local optimum, then these partial derivatives must necessarily vanish, that is,

$$\frac{\partial f}{\partial x_i} = \frac{\partial f}{\partial x_i} + \sum_{j=1}^{m} \frac{\partial g_i}{\partial h_j} \frac{\partial h_j}{\partial x_i} = 0, i = m+1, m+2, \cdots, n. \quad (6)$$

Now, if Equations (5) are used to substitute h_1, h_2, $\cdots$, h_m for x_1, x_2, $\cdots$, x_n, respectively, in Equation (1), then we obtain the identities

$$
\begin{aligned}
g_1 &= (h_1, h_2, \cdots, h_m, x_{m+1}, x_{m+2}, \cdots, x_n) \equiv 0 \\
g_2 &= (h_1, h_2, \cdots, h_m, x_{m+1}, x_{m+2}, \cdots, x_n) \equiv 0 \\
&\vdots \\
g_m &= (h_1, h_2, \cdots, h_m, x_{m+1}, x_{m+2}, \cdots, x_n) \equiv 0.
\end{aligned}
$$

By differentiating these identities with respect to x_{m+1}, x_{m+2}, $\cdots$, x_n we obtain

$$\frac{\partial g_k}{\partial x_i} + \sum_{j=1}^{m} \frac{\partial g_k}{\partial h_j} \frac{\partial h_j}{\partial x_i} = 0, \qquad (7)$$

$$i = m+1, m+2, \cdots, n; k = 1, 2, \cdots, m.$$

Let us now define the vectors

$$\boldsymbol{\delta}_k = \left(\frac{\partial g_k}{\partial x_{m+1}}, \frac{\partial g_k}{\partial x_{m+2}}, \cdots, \frac{\partial g_k}{\partial x_n} \right)', k = 1, 2, \cdots, m,$$

$$\boldsymbol{\gamma}_k = \left(\frac{\partial g_k}{\partial h_1}, \frac{\partial g_k}{\partial h_2}, \cdots, \frac{\partial g_k}{\partial h_m} \right)', k = 1, 2, \cdots, m,$$

$$\boldsymbol{\eta}_j = \left(\frac{\partial h_j}{\partial x_{m+1}}, \frac{\partial h_j}{\partial x_{m+2}}, \cdots, \frac{\partial h_j}{\partial x_n} \right)', j = 1, 2, \cdots, m,$$

$$\boldsymbol{\psi} = \left(\frac{\partial f}{\partial x_{m+1}}, \frac{\partial f}{\partial x_{m+2}}, \cdots, \frac{\partial f}{\partial x_n} \right)',$$

$$\boldsymbol{\tau} = \left(\frac{\partial f}{\partial h_1}, \frac{\partial f}{\partial h_2}, \cdots, \frac{\partial f}{\partial h_m} \right)'.$$

Equations (6) and (7) can then be written as

$$\left[\boldsymbol{\delta}_1 : \boldsymbol{\delta}_2 : \cdots : \boldsymbol{\delta}_m \right] + \left[\boldsymbol{\eta}_1 : \boldsymbol{\eta}_2 : \cdots : \boldsymbol{\eta}_m \right] \boldsymbol{\Gamma} = 0 \qquad (8)$$

$$\boldsymbol{\psi} + \left[\boldsymbol{\eta}_1 : \boldsymbol{\eta}_2 : \cdots : \boldsymbol{\eta}_m \right] \boldsymbol{\tau} = 0 \qquad (9)$$

where $\boldsymbol{\Gamma} = \left[\boldsymbol{\gamma}_1 : \boldsymbol{\gamma}_2 : \cdots : \boldsymbol{\gamma}_m \right]$, which is a nonsingular $m \times m$ matrix if condition (4) is valid.

From Equation (8) we have

$$\left[\boldsymbol{\eta}_1 : \boldsymbol{\eta}_2 : \cdots : \boldsymbol{\eta}_m \right] = -\left[\boldsymbol{\delta}_1 : \boldsymbol{\delta}_2 : \cdots : \boldsymbol{\delta}_m \right] \boldsymbol{\Gamma}^{-1}$$

By making the proper substitution in Equation (9) we obtain

$$\boldsymbol{\psi} + \left[\boldsymbol{\delta}_1 : \boldsymbol{\delta}_2 : \cdots : \boldsymbol{\delta}_m \right] \boldsymbol{\lambda} = 0 , \qquad (10)$$

where

$$\boldsymbol{\lambda} = -\boldsymbol{\Gamma}^{-1} \boldsymbol{\tau} . \qquad (11)$$

Equations (10) can then be expressed as

$$\frac{\partial f}{\partial x_i} + \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial x_i} = 0, i = m+1, m+2, \cdots, n. \quad (12)$$

From Equation (11) we also have

$$\frac{\partial f}{\partial x_i} + \sum_{j=1}^{m} \lambda_j \frac{\partial g_j}{\partial x_i} = 0, i = 1, 2, \cdots, m. \qquad (13)$$

Equations (12) and (13) can now be combined into a single vector equation of the form

$$\overline{V} f(x) + \sum_{j=1}^{m} \lambda_j \overline{V} g_j = 0 ,$$

which is the same as Equation (3). We conclude that at a stationary point of f, the values of x_1, x_2, $\cdots$, x_n and the corresponding values of λ_1, λ_2, $\cdots$, λ_m must satisfy Equations (1) and (3).

3. Lagrange Multipliers in Sampling Optimization

In two-phase sampling, or double sampling, from a population with size N we take one, relatively large, sample size n. From this relatively large sample we take a small sub-sample size m, which usually costs more per sample unit than the first one. In double sampling with regression estimators, the sample of the first phase n is used for the estimation of the average of an auxiliary variable X, $\overline{X}_1$, which should be strongly related to the main variable Y.

In the sub-sample m both auxiliary X and main Y variables are measured, in order to estimate their means $\overline{X}_2$ and $\overline{Y}_2$, respectively. The regression estimator $\hat{\overline{Y}}$ and its estimated variance $\mathrm{var}\,\hat{\overline{Y}}$ are ([7,11-13]):

$$\hat{\overline{Y}} = \overline{Y}_2 + b\left(\overline{X}_1 - \overline{X}_2 \right)$$

$$\mathrm{var}\,\hat{\bar{Y}} = s_{Y_m}^2 \left(\frac{1}{m} - r^2 \frac{n-m}{nm} \right)$$

where $s_{Y_m}^2$ is the variance of Y in the sub-sample m, and r is the estimated correlation coefficient between X and Y.

An approximate cost function could be

$$C = nC_1 + mC_2,$$

where

C: total sampling cost;
C_1: sampling cost of the first phase;
C_2: sampling cost of the second phase.

Sampling optimization can be achieved by minimizing cost C with fixed $\mathrm{var}\,\bar{Y}$, applying the following procedure:

We assume an approximately normal distribution of $\hat{\bar{Y}}$, so that the 95% confidence interval for $\bar{Y}$ would be:

$$\hat{\bar{Y}} - Z_{0.975}\sqrt{\mathrm{var}\,\hat{\bar{Y}}} < \bar{Y} < \hat{\bar{Y}} + Z_{0.975}\sqrt{\mathrm{var}\,\hat{\bar{Y}}},$$

where $0.975 = 1 - \alpha/2\,(\alpha = 1 - 0.95 = 0.05)$.

Now we must choose n and m in such a way, that half a confidence interval does not exceed a value D, fixed a priori, where D also may be expressed as a fraction (E) of $\mathrm{var}\,\hat{\bar{Y}}$:

$$Z_{0.975}\sqrt{\mathrm{var}\,\hat{\bar{Y}}} = E\hat{\bar{Y}} \Rightarrow \mathrm{var}\,\hat{\bar{Y}} = \frac{E^2\hat{\bar{Y}}^2}{Z_{0.975}^2}$$

To this end we construct the Lagrange function:

$$\varphi(n,m) = \mathrm{var}\,\hat{\bar{Y}}\left(\frac{1}{m} - r^2 \frac{n-m}{nm} \right) - \frac{E^2\hat{\bar{Y}}^2}{Z_{0.975}^2} = 0$$

$$F(n,m) = C + \lambda\varphi(n,m) = nC_1 + mC_2$$
$$+ \lambda\left[\mathrm{var}\,\hat{\bar{Y}}\left(\frac{1}{m} - r^2 \frac{n-m}{nm} \right) - \frac{E^2\hat{\bar{Y}}^2}{Z_{0.975}^2} \right]$$

Remind that

$$\left[f(x)/g(x) \right]' = \frac{f'(x)g(x) - f(x)g'(x)}{\left[g(x) \right]^2},$$

so

$$\frac{\partial F}{\partial n} = C_1 + \lambda\,\mathrm{var}\,\hat{\bar{Y}}\,\frac{-r^2\left[nm - (n-m)m \right]}{m^2 n^2}$$
$$= C_1 - \lambda\,\mathrm{var}\,\hat{\bar{Y}}\,\frac{r^2}{n^2} = 0 \tag{14}$$

$$\frac{\partial F}{\partial m} = C_2 + \lambda\,\mathrm{var}\,\hat{\bar{Y}}\left\{ \frac{-1}{m^2} - \frac{-r^2\left[-nm - (n-m)n \right]}{m^2 n^2} \right\}$$
$$= C_2 - \lambda\,\mathrm{var}\,\hat{\bar{Y}}\,\frac{(1-r^2)}{m^2} = 0 \tag{15}$$

$$\mathrm{var}\,\hat{\bar{Y}}\left(\frac{1}{m} - r^2 \frac{n-m}{nm} \right) - \frac{E^2\hat{\bar{Y}}^2}{Z_{0.975}^2} = 0 \tag{16}$$

Solving the system of Equations (14), (15) and (16) we find n, m and λ. The reverse problem, viz. finding $\mathrm{var}\,\bar{Y}$ for fixed C, is solved in a similar way.

In order to explain how Lagrange multipliers work, we describe the following example: Assume that the total cost C of an industry producing two products x and y, is given by the equation $C = 3x^2 + 6y^2 - xy$. The production is limited with a limitation of 20 units, that is $\lambda = x + y = 20$. Thus, we have:

$$C = F(x,y) = 3x^2 + 6y^2 - xy + \lambda(x + y - 20)$$

$$\frac{\partial F}{\partial x} = 6x - y + \lambda = 0 \tag{a}$$

$$\frac{\partial F}{\partial y} = 12y - x + \lambda = 0 \tag{b}$$

$$x + y = 20 \tag{c}$$

By solving the equations' system (a), (b) and (c), we find that $x = 13$, $y = 7$ and $\lambda = -71$. Consequently, $C = 3 \times 13^2 + 6 \times 7^2 - 13 \times 7 = 710$.

The economic meaning of λ is this: λ is the reduction of the total cost to the limit, if production was 19 instead of 20 units. In other words, if we required 19 total production units, the total cost would be reduced by 71 monetary units (710 − 71 = 639). Generally, λ represents the marginal effect on the cost function, when production limitation is increased by one unit.

4. Other Uses of Lagrange Multipliers in Forest Inventories

If there are not enough sample plots to give sufficiently good inventory results using only forest measurements, we may try to make use of auxiliary variables correlated with forest variables. The most obvious way is to use ratio or regression estimators (see Appendix 2). The calibration estimator of Deville and Särndal [14] is an extension of the regression estimator for obtaining population totals using auxiliary information. Both regression and calibration estimators can be employed if there are auxiliary variables for inventory sample plots known for which the population totals are also known, e.g. variables obtained from remote sensing or from GIS systems. The appeal of calibration estimators for forest inventories comes from the fact that they lead to estimators which are weighted sums of the sample plot variables, where the weight can be interpreted as the area of forest in the population that is similar to the sample plot.

The basic features of the calibration estimator of Deville and Särndal [14] in terms of estimating means can

be described as follows. Consider a finite population U consisting of N units. Let j denote a general unit, thus $U = \{1, \cdots, j, \cdots, N\}$. In a forest inventory the population is a region where units are pixels or potential sample plots. The units in a forest inventory will be referred to here as pixels, and it will be assumed that an inventory sample plot gives values to the forest variables for an associated pixel. Each unit j is associated with a variable yj and a vector of auxiliary variables xj. The population mean of $\boldsymbol{x}$, $\overline{\boldsymbol{X}} = N^{-1}\sum_U \boldsymbol{x}_j$ is assumed to be known. The y variables in a forest inventory are forest variables and the x variables can be spectral variables from remote sensing or geographical or climatic variables obtained from GIS databases.

Assume that a probability sample S is drawn, and y_j and $\boldsymbol{x}_j$ are observed for each j in S, the objective being to estimate the mean of y, $\overline{Y} = N^{-1}\sum_U y_j$. Let π_j be the inclusion probability and d_j the basic sampling design weight $d_j = (N\pi_j)^{-1}$, which can be used to compute the unbiased Horvitz-Thompson estimator

$$\hat{\overline{Y}}_d = \sum_S d_j y_j .$$

A calibration estimator

$$\hat{\overline{Y}} = \sum_S w_j y_j \tag{17}$$

is obtained by minimizing the sum of distances, $\sum_S G(w_j, d_j)$, between the prior weights d_j and posterior weights w_j for a positive distance function G, taking account of the calibration equation

$$\sum_S w_j \boldsymbol{x}_j = \overline{\boldsymbol{X}} . \tag{18}$$

If the distance between d_j and w_j is defined as

$$G_1 = (w_j, d_j) = (w_j - d_j)^2 / d_j ,$$

the calibration estimator will be the same as the regression estimator

$$\hat{\overline{Y}}_r = \sum_S w_j y_j = \hat{\overline{Y}}_d + (\overline{\boldsymbol{X}} - \overline{\boldsymbol{X}}_d)' \hat{\boldsymbol{b}} , \tag{19}$$

where $\hat{\overline{\boldsymbol{X}}}_d$ and $\hat{\boldsymbol{b}}$ (a weighted regression coefficient vector) are

$$\hat{\overline{\boldsymbol{X}}}_d = \sum_S d_j \boldsymbol{x}_j \tag{20}$$

and

$$\hat{\boldsymbol{b}} = \left(\sum_S d_j \boldsymbol{x}_j \boldsymbol{x}_j'\right)^{-1} \sum_S d_j \boldsymbol{x}_j y_j . \tag{21}$$

If the model contains an intercept, the corresponding variable x will be one for all observations, and the calibration Equation (18) will then guarantee that the weights w_j add up to one. This means that when estimating totals, the weights Nw_j will add up to the known total number of pixels in the population. Thus Nw_j can be in-

terpreted as the total area, in pixel units, for plots of forest similar to plot j. The standard least squares theory implies that the regression estimator (19) can be expressed in the form

$$\hat{\overline{Y}}_r = \sum_S w_j y_j = \overline{\boldsymbol{X}}' \hat{\boldsymbol{b}} . \tag{22}$$

It is assumed that the intercept is always among the parameters. Estimator (21) is defined if the moment matrix $\sum_S d_j \boldsymbol{x}_j \boldsymbol{x}_j'$ is non-singular.

Some of the weights w_j in (17) implied by Equations (20)-(22) may be negative. Nonnegative weights are guaranteed if the distance function is infinite for negative w_j. Deville and Särndal [14] presented four distance functions producing positive weights.

Minimization of the sum $\sum_S G(w_j, d_j)$, so that (18) is satisfied is a non-linear constrained minimization problem. Using Lagrange multipliers, the problem can be reformulated as a non-linear system of equations which can be solved iteratively using Newton's method [14]. If the initial values of the Lagrange multipliers are set to zero, the first step will produce w_j's of the regression estimator (19).

Since the calibration estimator is asymptotically equivalent to the regression estimator, Deville and Särndal [14] suggest that the variance of the calibration estimator should be computed in the same way as the variance of the regression estimator using regression residuals. There is no design-unbiased estimator of the variance in systematic sampling [7].

The emphasis on area interpretation for the weights has the same argument behind it as was used by Moeur and Stage [14] for the most similar neighbour method (MSN), where unknown plot variables are taken from a plot which is as similar as possible with respect to the known plot variables. In both methods each sample plot represents a percentage of the total area, and all the forest variables are logically related to each other. The difference is that in the calibration estimator we obtain an estimate of the area of the sample plot for the whole population whereas in the MSN method each pixel is associated with a sample plot. Since there is no straightforward way of showing that the MSN method produces optimal results in any way at the population level, it may be safer to use the calibration estimator for computing population-level estimates for forest variables. The problem with the calibration estimator is that it does not provide a map. If a map is needed, then the weights provided by the calibration estimator need to be distributed over pixels using separate after-processing.

Lappi [15] proposed a small-area modification of the calibration estimator which can be used when several subpopulation totals are required simultaneously. He used satellite data as auxiliary information for computing

inventory results for counties. Sample plots in the surrounding inclusion zone are also used for a given subpopulation so that the prior weight decreases as distance increases. The error variance is computed using a spatial variogram model. Block kriging [16] provides an optimal estimator for subpopulation totals under such a model, but kriging can produce negative weights for sample plots, and the weights are different for each y variable. Thus it is not possible to give areal interpretations to sample plot weights in kriging.

REFERENCES

[1] J. Penman, M. Gytarsky, T. Hiraishi, T. Krug, D. Kruger, R. Pipatti, L. Buendia, K. Miwa, T. Ngara, K. Tanabe and F. Wagner, "Good Practice Guidance for Land Use, Land-Use Change and Forestry," *Intergovernmental Panel on Climate Change*: *IPCC National Greenhouse Gas Inventories Program*, Institute for Global Environmental Strategies (IGES) for the IPCC, Kanagawa Japan, 2003.

[2] F. Loetsch, F. Zöhrer and K. Haller, "Forest Inventory," BLV, Verlagsgesellschaft, 1973.

[3] I. Doig, "When the Douglas-Firs Were Counted: The Beginning of the Forest Survey," *Journal of Forest History*, Vol. 20, 1976, pp. 20-27.

[4] W. Frayer and G. Furnival, "Forest Survey Sampling Designs—A History," *Journal of Forestry*, Vol. 97, No. 12, 1999, pp. 4-10.

[5] T. Gregoire, "Roots of Forest Inventory in North America," A Paper Presented at the A1 Inventory Working Group Session at the SAF National Convention Held at Richmond, VA, 25-28 October 1992.

[6] T. Honer and F. Hegyi, "Forest Inventory—Growth and Yield in Canada: Past, Present and Future," *The Forestry Chronicle*, Vol. 66, 1990, pp. 112-117.

[7] H. Schreuder, T. Gregoire and G. Wood, "Sampling Methods for Multiresource Forest Inventory," John Wiley and Sons, New York, 1993.

[8] R. Seppälä, "Forest Inventories and the Development of Sampling Methods," *Silva Fennica*, Vol. 3, 1985, pp. 218-219.

[9] D. Van Hooser, N. Cost and H. Lund, "The History of Forest Survey Program in the United States," In: G. Preto and B. Koch, Eds., *Forest Resource Inventory and Monitoring and Remote Sensing Technology*, *Proceedings of the IUFRO Centennial Meeting in Berlin*, 31 August-4 September 1992, Japan Society of Forest Planning Press, Tokyo University of Agriculture and Technology, Saiwaicho, Fuchu, Tokyo, 1992, pp. 19-27.

[10] S. Rao, "Optimization Theory and Application," Wiley Eastern Ltd., New Delhi, 1984.

[11] W. Cochran, "Sampling Techniques—3rd Edition," Wiley, New York, 1977.

[12] P. De Vries, "Sampling for Forest Inventory," Springer, Berlin, 1986. doi:10.1007/978-3-642-71581-5

[13] K. Matis, "Sampling of Natural Resources," Pigasos Publications, Thessaloniki, Greece, 2004.

[14] J. Deville and C. Särndal, "Calibration Estimators in Survey Sampling," *Journal of the American Statistical Association*, Vol. 87, No. 418, 1992, pp. 376-382. doi:10.2307/2290268

[15] J. Lappi, "Forest Inventory of Small Areas Combining the Calibration Estimator and a Spatial Model," *Canadian Journal of Forest Research*, Vol. 31, No. 9, 2001, pp. 1551-1560. doi:10.1139/x01-078

[16] N. Cressie, "Kriging Nonstationary Data," *Journal of American Statistical Association*, Vol. 81, No. 395, 1986, pp. 625-634. doi:10.2307/2288990

[17] M. Moeur and A. Stage, "Most Similar Neighbor: An Improved Sampling Inference Procedure for Natural Resource Planning," *Forest Science*, Vol. 41, No. 2, 1995, pp. 337-359.

Appendices

Appendix 1. Implicit Function Theorem

Let $g : D \to R^n$, where D is an open subset of R^{m+n}, and g has continuous first-order partial derivatives in D. If there is a point $z_0 \in D$, where $z_0 = (x_0', y_0')'$, with $x_0 \in R^n$, $y_0 \in R^m$ such that $g(z_0) = 0$, and if at z_0,

$$\frac{\partial (g_1, g_2, \cdots, g_n)}{\partial (x_1, x_2, \cdots, x_n)} \neq 0,$$

where g_i is the ith element of g $(i = 1, 2, \cdots n)$, then there is a neighborhood $N_\delta(y_0)$ of y_0 in which the equation $g(x, y) = 0$ can be solved uniquely for x as a continuously differentiable function of y.

Appendix 2. Ratio and Regression Estimators

In a stratified inventory information on some auxiliary variables is used both to plan the sampling design (e.g. allocation) and for estimation, or only for estimation (post-stratification). Stratification is not the only way to use auxiliary information, however, as it can be used at the design stage, e.g. in sampling proportional to size (see Appendix 3). It can also be used at the estimation stage in ratio or regression estimators, so that the standard error of the estimators can be reduced using information on a variable x which is known for each sampling unit in the population. The estimation is based on the relationship between the variables x and y. In ratio estimation, a model that goes through the origin is applied. If this model does not apply, regression estimator is more suitable. The ratio estimator for the mean is

$$\hat{\bar{y}}_{rat} = \frac{\bar{y}}{\bar{x}} \bar{X} = r\bar{X},$$

where $\bar{X}$ is the mean of a variable x in the population and $\bar{x}$ in the sample. Ratio estimators are usually biased, and thus the root mean square error (RMSE) should be used instead of the standard error. The relative bias nevertheless decreases as a function of sample size, so that in large samples (at least more than 30 units) the accuracy of the mean estimator can be approximated as [1]:

$$\mathrm{var}\left(\hat{\bar{y}}_{rat}\right) \cong \left(1 - \frac{n}{N}\right) \sum_{i=1}^{n} \frac{(y_i - rx_i)^2}{n(n-1)}.$$

The ratio estimator is more efficient the larger the correlation between x and y relative to the ratio of the coefficients of variation CV. It is worthwhile using the ratio estimator if

$$corr(x, y) > \frac{1}{2} \frac{CV(x)}{CV(y)}.$$

The (simple linear) regression estimator for the mean value is

$$\hat{\bar{y}}_{reg} = \bar{y} + \hat{\beta}\left(\bar{X} - \bar{x}\right),$$

where $\hat{\beta}$ is the OLS (Ordinary Least Squares) coefficient of x for the model, which predicts the population mean of y based on the sample means. In a sampling context, the constant of the model is not usually presented, but the formula for the constant, $\hat{\alpha} = \bar{y} + \hat{\beta}\bar{x}$, is embedded in the equation. The model is more efficient the larger the correlation between x and y. The variance of the regression estimator can be estimated as

$$\mathrm{var}\left(\hat{\bar{y}}_{reg}\right) \cong \left(1 - \frac{n}{N}\right) \sum_{i=1}^{n} \frac{\left[(y_i - \bar{y}) - \hat{\beta}(x_i - \bar{x})\right]^2}{n(n-2)}.$$

Appendix 3. Sampling with Probability Proportional to Size

The basic properties of sampling with arbitrary probabilities can also be utilized in sampling with probability proportional to size (PPS), such as sampling with a relascope. It is then assumed that unit i is selected with the probability kx_i, where k is a constant and x is a covariate (diameter of a tree in relascope sampling). PPS sampling is more efficient the larger the correlation between x and y. For perfect correlation the variance in the estimator would be zero [7]. PPS sampling might even be less efficient than SRS (Simple Random Sampling), however, if the correlation were negative. This could be the case when multiple variables of interest are considered simultaneously, for example, when correlation with one variable (say volume) might give efficient estimates but the estimates for other variables (say health and quality) might not be so good.

In practice, PPS sampling can be performed by ordering the units, calculating the sum of their sizes (say $\sum x_i$), and calculating $\sum x_i / n$. The probability of a unit i being selected is then $x_i / \sum x_i$ and a cumulative probability can be calculated for the ordered units. A random number r is then picked and each unit with a cumulative probability equal to (or just above) r, $r + 1$, $r + 2$, $\cdots$, $r + n - 1$ is selected for the sample. Every unit of size greater than $\sum x_i / n$ is then selected with certainty.